arXiv:1301.6371v1 [math.CO] 27 Jan 2013

Shattering Thresholds for Random Systems of Sets, Words, and Permutations

Anant P. Godbole
Department of Mathematics and Statistics
East Tennessee State University

Samantha Pinella
School of Mathematics
University of Edinburgh

Yan Zhuang
Department of Mathematics and Computer Science
Goucher College

January 29, 2013

Abstract

This paper considers a problem that relates to the theories of covering arrays [5], permutation patterns [8], Vapnik-Červonenkis (VC) classes [2],[6], and probability thresholds [1]. Specifically, we want to find the number of subsets of $[n] := \{1, 2, \ldots, n\}$ we need to randomly select, in a certain probability space, so as to respectively "shatter" all t-subsets of $[n]$. Moving from subsets to words, we ask for the number of n-letter words on a q-letter alphabet that are needed to shatter all t-subwords of the q^n words of length n. Finally, we explore the number of permutations of $[n]$ needed to shatter (specializing to $t = 3$), all length 3 permutation patterns in specified positions. We uncover a very sharp zero-one probability threshold for the emergence of such shattering; Talagrand's isoperimetric inequality in product spaces [1] is used as a key tool.

1 Introduction

In this section, we give the necessary background on covering arrays; Vapnik-Červonenkis classes; and permutation patterns, and then explain our goals. A $k \times n$ array with entries from the alphabet $\{0, 1, \ldots, q-1\}$ is said to be a (t, q, n, k, λ)-covering array, or briefly a t-covering array, if for each of the $\binom{n}{t}$ choices of t columns, each of the q^t q-ary words of length t can be found at least λ times among the rows of the selected columns. Covering arrays are used as valuable tools in software testing; see, e.g., [5], which is a comprehensive survey of the theory of t-covering arrays. In this paper, we will focus solely on the case $\lambda = 1$. If $q = 2$, we can interpret any row as the characteristic vector of a subset of $[n]$ – by making a correspondence between the positions where the row has ones, and the set of those positions. We thus have the following alternative formulation of covering arrays: A family $\mathcal{F}$ of subsets of $[n]$ is a t-covering array if for each $\{a_1, \ldots, a_t\} \subset [n]$,

$$|\{\{a_1, \ldots, a_t\} \cap F\} : F \in \mathcal{F}| = 2^t.$$

We next see how this definition relates to that of VC classes.

A class $\mathcal{F}$ of subsets of an abstract set $\mathcal{Y}$ is said to *shatter* a subset $A = \{a_1, \ldots, a_t\} \subseteq \mathcal{Y}$ if

$$|\{A \cap F\} : F \in \mathcal{F}| = 2^t.$$

Furthermore, the VC dimension $\mathcal{VC}(\mathcal{F})$ of $\mathcal{F}$ [6] is the cardinality of the smallest *unshattered* set (the dimension is ∞ if all sets of all finite size are shattered.) A class $\mathcal{F}$ is said to be a VC class if $\mathcal{VC}(\mathcal{F}) < \infty$. Many canonical examples of VC classes are driven by underlying geometric considerations. For example, consider the *infinite* family $\mathcal{F}$ of subsets of $\mathcal{Y} = \mathbb{R}$ of the form $(-\infty, x] : x \in \mathbb{R}$. Every set of size 1 is clearly shattered by $\mathcal{F}$. Next letting $t = 2$, we consider the class of all 2-element subsets of $\mathbb{R}$. It is then clear that

$$|\{\{a_1, a_2; a_1 < a_2\} \cap F\} : F \in \mathcal{F}| = 3 < 2^2,$$

since it is impossible for $F \in \mathcal{F}$ to intersect the two element set $\{a_1, a_2\}$ in its larger element a_2. It follows that $\mathcal{VC}(\mathcal{F}) = 2$. To give another example, if $\mathcal{F}$ consists of all convex sets in $\mathcal{Y} = \mathbb{R}^2$, then it is impossible for elements of $\mathcal{F}$ to "shatter" a three element subset $A = \{a_1, a_2, a_3\}$ of collinear points since

$$|\{A \cap F : F \in \mathcal{F}\}| = 7 < 2^3.$$

Since every 2-element set can be shattered by convex sets, we have that $\mathcal{VC}(\mathcal{F}) = 3$ in this case.

VC classes were first defined and used in the context of uniform limit theorems in Statistics [6], [13]; later, their use was extended to learning theory [2], [12]. The alternative (and perhaps more popular) definition of the VC dimension of $\mathcal{F}$ is "the cardinality of the largest shattered set". In many cases, e.g., in the first example given above, the largest shattered set and the smallest unshattered set differ in size by 1; in general, however, this is not the case, as in the second example. We will use the first definition of the VC dimension for reasons that will become clear in what follows, but which all stem from the fact that we are operating in a finite setting.

The above discussion reveals that with $\mathcal{Y} = [n]$, an ensemble $\mathcal{F}$ of finite subsets of $[n]$ is a binary t-covering array if and only if $\mathcal{VC}(\mathcal{F}) \geq t+1$; an explanation follows: If $\mathcal{F}$ is t-covering, then for each set A of size t and each $B \subseteq A$, there exists $F \in \mathcal{F}$ such that $F \cap A = B$; thus every set of size t is shattered, and the smallest unshattered set must be of size $t+1$ or more. The reverse argument is valid too.

When $q \geq 3$, covering arrays are often described in terms of words. We will use this terminology in this paper too, but the notions of shattering and VC dimension are probably best described using the language of multisets. We will interpret a $k \times n$ array $\{a_{ij}\}_{1 \leq i \leq k; 1 \leq j \leq n}$, with entries from $\{0, 1, \ldots, q-1\}$ as consisting of k multisets, with the ith multiset containing the element j a_{ij} times, where the *degree* of the multiset, i.e., the maximum number of times an element may appear in it, is bounded by $q-1$. The notion of the intersection of two multisets A, B is defined in the natural way; for example $\{1,1,2,2,3\} \cap \{1,1,1,1,2,3,3\} = \{1,1,2,3\}$. We say that a collection $\mathcal{F}$ of k multisets shatters a multiset A with t distinct elements each repeated $q-1$ times, if

$$|\{A \cap F : F \in \mathcal{F}\}| = q^t.$$

As before the VC dimension $\mathcal{VC}(\mathcal{F})$ of $\mathcal{F}$ is the cardinality of the smallest *unshattered* multiset of the above type, with q fixed and the minimum taken over t ($\mathcal{VC}(\mathcal{F}) = \infty$ if there is no such smallest t). We thus see that $\mathcal{F}$ is a $(t, q, n, k, 1)$-covering array if and only if $\mathcal{VC}(\mathcal{F}) \geq t+1$.

We next turn to permutations. The theory of permutation patterns was initiated by Knuth [8], and continues to be an area of active investigation.

We say that a permutation $\pi \in S_n$ *contains* the permutation $\rho \in S_t$ if there exist indices $1 \le i_1 < i_2 < \ldots < i_t \le n$ such that $(\pi_{i_1}, \ldots, \pi_{i_t})$ and $(\rho_1, \ldots, \rho_t)$ are order isomorphic; if not we say that π *avoids* ρ. Enumeration questions are critical in this area; for example it is known that for $t = 3$, the number of (i, j, k) avoiding n-permutations is given by the Catalan numbers $\binom{2n}{n}/(n+1)$ for each of the six choices of i, j, k [3]. Moreover, the Stanley-Wilf conjecture, namely that for fixed t, the number of ρ-avoiding n-permutations is asymptotic to C^n for some $1 \le C = C_\rho < \infty$, was recently proved by Marcus and Tardos [10]. How might shattering and VC dimension be defined in the case that $\mathcal{F}$ consists of an array of k n-permutations $(\pi_1, \ldots, \pi_k)$? Using the language of covering arrays, we shall say that the VC dimension is at least $t+1$ if for each choice of t columns and $\rho \in S_t$, at least one row of the selected columns contains entries order isomorphic to those of ρ. This is equivalent to saying that the k permutations, restricted to any t positions, shatter all the $t!$ permutations on those positions.

Research on t-covering arrays has focused on finding arrays of small size. In [11], for example, the case of $t = 3$ is studied in detail, and Roux's result that there exist 3-covering binary arrays of size $7.5\mathrm{lg}n$ is proved, where $\mathrm{lg} = \log_2$. This result was re-proved in [7] using the Lovász Local Lemma (see [1]), where the underlying probability model consisted, as in the work of Roux, of independently placing an equal number of ones and zeros in each column. This model is intractable for general values of t and q; accordingly, the general upper bound on the size of covering arrays was proved in [7] by reverting to a simple multinomial model, where each spot in the $k \times n$ array is independently and uniformly chosen from the set $\{0, 1, \ldots, q-1\}$. But the Lovász Lemma is an existence result whose conclusion is that there is a positive probability that there are no "bad events," i.e., that

$$k \ge K \Rightarrow \mathbb{P}(\text{array is } t-\text{covering}) > 0,$$

so that a t-covering array with K rows exists. By contrast, in this paper we are looking for results, still in the $\log n$ domain, that are of the form

$$k \le k_0(n) \Rightarrow \mathbb{P}(\text{array is } t-\text{covering}) \to 0 \ \ (n \to \infty);$$

$$k \ge k_1(n) \Rightarrow \mathbb{P}(\text{array is } t-\text{covering}) \to 1 \ \ (n \to \infty),$$

and where the gap $[k_0(n), k_1(n)]$ is not too wide. We will use the simple first moment method (linearity of expectation) together with Talagrand's isoperimetric inequalities, to establish such a result in Section 2.

The situation is a little more nuanced when we turn to the question of shattering permutations. First of all, we are only able to prove clean results when $t = 3$, but, more importantly, it is also meaningful to consider *large* arrays with small VC dimension. For example, if we wrote each of the $\sim 4^n$ 123-avoiding n permutations in a stack, there would be no 3-ple order isomorphic to 123 in *any* set of 3 columns, and the relevant question would be to investigate how much better than that we could do. This is the approach taken by Cibulka and Kyncl [4], who, use the second definition of VC dimension to give superexponential bounds on the size of the stack for $t = 3$. Our motivation is to use the first definition of VC definition to produce logarithmic stacks with given VC dimension – with probability that transitions from 0 to 1 in a short interval. Results along these lines are proved in Section 3.

2 Shattering Subsets and Words

We use the following model. Let $\mathcal{F}$ be a randomly generated stack of k words, each of length n and obtained by selecting each position in the $k \times n$ array to independently and uniformly be one of the letters of the "alphabet" $\{0, 1, \ldots, q-1\}$. Denote the words in $\mathcal{F}$ as $F_1, \ldots, F_k$. As noted in Section 1, if $q = 2$, then $\mathcal{F}$ is simply a random set system of k subsets of $[n]$. We will use *rows* to refer to the words in $\mathcal{F}$ and *columns* to refer to the character positions. In this section, we show that the threshold, under our model, for the property "$\mathcal{F}$ shatters all t-words" (which is an alternative term we use for multiset shattering) occurs at the level $\frac{t}{\lg\left(\frac{q^t}{q^t-1}\right)}\lg(n)$; this will allow us to determine with high probability the VC dimension of a random word array. Deriving an upper threshold is easy:

Theorem 2.1. *If $k \geq \frac{t \lg n}{\lg\left(\frac{q^t}{q^t-1}\right)}(1+o(1))$, then all t-words are shattered almost surely by $\mathcal{F}$, i.e., the probability that $\mathcal{F}$ is a covering array tends to 1 as $n \to \infty$.*

Proof. Let X be the number of sets of t columns corresponding to unshattered t-words. By Markov's inequality, $\mathbb{P}(X \geq a) \leq \mathbb{E}(X)/a$, valid for non-negative random variables X, we have:

$$\mathbb{P}\left(X \geq 1\right) \leq \mathbb{E}\left(X\right) \leq \binom{n}{t} q^t \left(\frac{q^t - 1}{q^t}\right)^k \leq \frac{n^t}{t!} q^t \left(\frac{q^t - 1}{q^t}\right)^k \to 0$$

provided that (with $\omega(n)$ denoting a function growing to infinity arbitrarily slowly)

$$k \geq \frac{t\lg n + \omega(n) - \lg t! + t\lg q}{\lg\left(\frac{q^t}{q^t-1}\right)} = \frac{t\lg n}{\lg\left(\frac{q^t}{q^t-1}\right)}(1 + o(1)) := k_1(n),$$

as asserted. □

Proving that the lower threshold function $k_0(n)$ is of the same magnitude is tantamount to showing that the random variable is sharply concentrated around its mean. In some sense this was done in [7], but using a naïve (and ultimately incorrect) probability model that was only shown to be valid for $t = 3$. To give a more rigorous proof in this paper, we shall apply Talagrand's inequality in the form found in [1] (other proofs are possible). This inequality is applicable for random variables that are *1-Lipschitz*:

Definition 2.2. *Let Z be a random variable expressed as a function of N independent indicator variables I_i. We call Z 1-Lipschitz if*

$$|Z\left(I_1, \ldots, I_N\right) - Z\left(I_1^*, \ldots, I_N^*\right)| \leq 1$$

whenever $I_i \neq I_i^*$ for at most one i.

The random variable X, counting the number of sets of "defective" t columns (i.e. those sets corresponding to unshattered t-words), depends on nk mutually independent random variables. It is not, however 1-Lipschitz since an added presence or absence of a specific character may change X by more than 1 due to overlapping columns. However, if we define Y as the maximum number of *non-overlapping* sets of "defective" columns, then Y is 1-Lipschitz.

Talagrand's inequality also involves the notion of a *certification function*:

Definition 2.3. *Let Z be a random variable expressed as a function of N indicator variables I_i, and let $f : \mathbb{N} \to \mathbb{N}$ be a function. We call Z f-certifiable if $Z \geq s$ can always be verified to be true by $f(s)$ of the N indicator variables. In this case, we call f a certification function for Z.*

Given the random variable Y as above, to verify that there are at least s non-overlapping sets of unshattered k-words, it is easy to see that it suffices to know kts of the entries in the array. Thus f is linear and $f(s) = kts$.

Talagrand's inequality is reproduced below for completeness:

Theorem 2.4 (**Talagrand's Inequality**). *Let Z be a 1-Lipschitz random variable with certification function f. Then, for all $m, u > 0$:*

$$\mathbb{P}(Z \leq m - u\sqrt{f(m)})\mathbb{P}(Z \geq m) \leq e^{-\frac{u^2}{4}}$$

Applying Talagrand's inequality to the variable Y with $m = \mathrm{Med}(Y)$ (so that $\mathbb{P}(Y \geq m) \geq 1/2$) and $u = \sqrt{\frac{m}{kt}}$, we see that

$$\mathbb{P}(Y = 0) \leq 2e^{-\frac{m}{4kt}}. \tag{1}$$

We will use (1) in an appropriate way to get the lower threshold; specifically, we need to derive conditions under which (i) $\mathbb{E}(Y)$ and $\mathbb{E}(X)$ are close; and (ii) $\mathbb{E}(Y)$ and m are close. A series of technical lemmas that lead to (i) and (ii) are presented next.

Lemma 2.5. *Let Γ and Δ be distinct non-disjoint sets of t columns. Let r be the number of overlapping elements of Γ and Δ; i.e., $r = |\Gamma \cap \Delta|$. Define the indicator random variable I_Γ as being 1 if Γ is missing at least one t-word and 0 otherwise. Then*

$$\mathbb{P}\left(I_\Gamma I_\Delta = 1\right) \leq q^{2t-r}\left(\frac{q^t + q^{r-t} - 2}{q^t}\right)^k \{1 + o(1)\} \quad (k \to \infty).$$

Proof. Lemma 2.5 generalizes a result in [7] and this proof is similar. Let A_m be the event that exactly m words are missing from Γ. We have

$$
\begin{aligned}
\mathbb{P}\left(I_\Gamma I_\Delta = 1\right) &= \mathbb{P}\left(I_\Gamma = 1\right)\mathbb{P}\left(I_\Delta = 1 | I_\Gamma = 1\right) \\
&= \mathbb{P}\left(I_\Gamma = 1\right)\mathbb{P}\left(I_\Delta = 1 | A_1 \cup A_2 \cup \cdots \cup A_{q^t-1}\right) \\
&= \mathbb{P}\left(I_\Gamma = 1\right)\left[\frac{\mathbb{P}\left(I_\Delta = 1 \cap A_1\right) + \cdots + \mathbb{P}\left(I_\Delta = 1 \cap A_{q^t-1}\right)}{\mathbb{P}\left(A_1 \cup A_2 \cup \ldots \cup A_{q^t-1}\right)\right)}\right] \\
&\leq \mathbb{P}\left(I_\Gamma = 1\right)\left[\mathbb{P}\left(I_\Delta = 1 | A_1\right) + \left(\frac{\mathbb{P}\left(A_2\right) + \cdots + \mathbb{P}\left(A_{q^t-1}\right)}{\mathbb{P}\left(A_1 \cup \ldots \cup A_{q^t-1}\right)}\right)\right] \\
&\leq \mathbb{P}\left(I_\Gamma = 1\right)\left[\mathbb{P}\left(I_\Delta = 1 | A_1\right) + \frac{\binom{q^t}{2}\left(\frac{q^t-2}{q^t}\right)^k}{q^t\left(\frac{q^t-1}{q^t}\right)^k - \binom{q^t}{2}\left(\frac{q^t-2}{q^t}\right)^k}\right] \\
&= \mathbb{P}\left(I_\Gamma = 1\right) \cdot \\
&\quad \left[\mathbb{P}\left(I_\Delta = 1 | A_1\right) + \frac{\binom{q^t}{2}\left(\frac{q^t-2}{q^t}\right)^k}{q^t\left(\frac{q^t-1}{q^t}\right)^k\left(1 - \left(q^t-1\right)\left(\frac{q^t-2}{q^t-1}\right)^k\right)}\right] \\
&= \mathbb{P}\left(I_\Gamma = 1\right) \cdot \\
&\quad \left[\mathbb{P}\left(I_\Delta = 1 | A_1\right) + \left(\frac{q^t-1}{2}\right)\left(\frac{q^t-2}{q^t-1}\right)^k \{1 + o(1)\}\right]. \quad (2)
\end{aligned}
$$

Since $\mathbb{P}\left(I_\Gamma = 1\right) \leq q^t\left(1 - q^{-t}\right)^k$, the problem reduces to upper-bounding $\mathbb{P}\left(I_\Delta = 1 | A_1\right)$. Exactly one word is missing in Γ; let us denote that word by γ. Assume, without loss of generality, that the first r columns of Δ are the same as the last r columns of Γ. We consider two cases. Let p_1 be the conditional probability that a word beginning with the last r characters of γ is also missing in Δ; there are q^{t-r} such words. Let p_2 be the probability that a word not beginning with these same r characters is missing in Δ; there are $q^t - q^{t-r}$ such words. Hence:

$$
\mathbb{P}\left(I_\Delta = 1 | A_1\right) \leq q^{t-r} p_1 + \left(q^t - q^{t-r}\right) p_2
$$

We first calculate p_1. We know that γ is the only word missing in Γ, so for each of the remaining $q^t - 1$ words in this first category, there exists a row in Γ containing that word. Take away one such row for each of these $q^t - 1$ words; each of the other rows are randomly assigned to one of these

$q^t - 1$ words with probability $\frac{1}{q^t-1}$ each. This process enables one to realize the probability distribution of the content of the rows of Γ given that A_1 has occurred. Let $\mathcal{A}$ be the number of rows in Δ that coincide with those of Γ in the overlapping r positions; note that $\mathcal{A}$ is at least $q^{t-r} - 1$. Then for $a \geq 0$,

$$\mathbb{P}(\mathcal{A} = a + q^{t-r} - 1) = \binom{k - (q^t - 1)}{a} \left(\frac{q^{t-r} - 1}{q^t - 1}\right)^a \left(\frac{q^t - q^{t-r}}{q^t - 1}\right)^{k - \left(q^t - 1\right) - a}$$

Using the binomial theorem we obtain

$$\begin{aligned} p_1 &= \sum_{a=0}^{k-\left(q^t-1\right)} \binom{k - (q^t - 1)}{a} \left(\frac{q^{t-r} - 1}{q^t - 1}\right)^a \left(\frac{q^t - q^{t-r}}{q^t - 1}\right)^{k-\left(q^t-1\right)-a} \cdot \\ &\qquad \left(1 - \frac{1}{q^{t-r}}\right)^{a+q^{t-r}-1} \\ &= \left(\frac{\left(q^{t-r} - 1\right)^2}{q^{t-r}\left(q^t - 1\right)} + \frac{q^t - q^{t-r}}{q^t - 1}\right)^{k-\left(q^t-1\right)} \left(\frac{q^{t-r} - 1}{q^{t-r}}\right)^{q^{t-r}-1} \\ &= \left(1 - \frac{1 - q^{r-t}}{q^t - 1}\right)^k \left(1 - \frac{1 - q^{r-t}}{q^t - 1}\right)^{-\left(q^t-1\right)} \left(1 - \frac{1}{q^{t-r}}\right)^{q^{t-r}-1} \\ &\leq \left(1 - \frac{1 - q^{r-t}}{q^t - 1}\right)^k, \end{aligned}$$

where the last inequality is valid since

$$\left(1 - \frac{1}{q^{t-r}}\right) \leq \left(1 - \frac{1}{q^t - 1} + \frac{1}{q^{t-r}(q^t - 1)}\right)^{\frac{q^t-1}{q^{t-r}-1}},$$

which follows from the fact that the function $(1 - kx)^{1/x}$ is monotone decreasing on the interval $[0, 1]$ for fixed $k \in (0, 1)$.

Repeating the process for p_2, let $\mathcal{B}$ be the number of rows in Δ that do not begin with the last r characters of γ in some fixed fashion; $\mathcal{B}$ is at least q^{t-r}. Then,

$$\mathbb{P}(\mathcal{B}=b+q^{t-r})=\binom{k-(q^t-1)}{b}\left(\frac{q^{t-r}}{q^t-1}\right)^b\left(\frac{q^t-1-q^{t-r}}{q^t-1}\right)^{k-(q^t-1)-b},$$

and by the same reasoning as before,

$$\begin{aligned}
p_2 = & \sum_{b=0}^{k-\left(q^t-1\right)}\binom{k-(q^t-1)}{b}\left(\frac{q^{t-r}}{q^t-1}\right)^b\left(\frac{q^t-1-q^{t-r}}{q^t-1}\right)^{k-\left(q^t-1\right)-b}\cdot \\
& \qquad \left(1-\frac{1}{q^{t-r}}\right)^{b+q^{t-r}} \\
= & \left(1-\frac{1}{q^t-1}\right)^k\left(1-\frac{1}{q^t-1}\right)^{-\left(q^t-1\right)}\left(1-\frac{1}{q^{t-r}}\right)^{q^{t-r}} \\
\leq & \left(1-\frac{1}{q^t-1}\right)^k.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{P}\left(I_\Delta=1|A_1\right) &\leq q^{t-r}p_1+\left(q^t-q^{t-r}\right)p_2 \\
&\leq q^{t-r}\left(1-\frac{1-q^{r-t}}{q^t-1}\right)^k+\left(q^t-q^{t-r}\right)\left(1-\frac{1}{q^t-1}\right)^k \\
&= q^{t-r}\left(1-\frac{1-q^{r-t}}{q^t-1}\right)^k\left(1+(q^r-1)\left(\frac{q^t-2}{q^{r-t}+q^t-2}\right)^k\right) \\
&= q^{t-r}\left(1-\frac{1-q^{r-t}}{q^t-1}\right)^k\{1+o\left(1\right)\}\quad(k\to\infty), \qquad (3)
\end{aligned}$$

and thus by (2) and (3)

$$\begin{aligned}
\mathbb{P}\left(I_\Gamma I_\Delta = 1\right) &\le \mathbb{P}\left(I_\Gamma = 1\right)\left[\mathbb{P}\left(I_\Delta = 1|A_1\right) + \left(\frac{q^t-1}{2}\right)\left(\frac{q^t-2}{q^t-1}\right)^k\{1+o\left(1\right)\}\right] \\
&\le q^t\left(1-q^{-t}\right)^k \cdot \\
&\quad \left[q^{t-r}\left(1-\frac{1-q^{r-t}}{q^t-1}\right)^k + \left(\frac{q^t-1}{2}\right)\left(\frac{q^t-2}{q^t-1}\right)^k\right]\{1+o(1)\} \\
&= \left(q^{2t-r}\left(\frac{q^t+q^{r-t}-2}{q^t}\right)^k + \frac{q^t\left(q^t-1\right)}{2}\left(\frac{q^t-2}{q^t}\right)^k\right)\{1+o\left(1\right)\} \\
&= q^{2t-r}\left(\frac{q^t+q^{r-t}-2}{q^t}\right)^k \cdot \\
&\quad \left(1+\frac{q^t-1}{2q^{t-r}}\left(\frac{q^t-2}{q^t+q^{r-t}-2}\right)^k\right)\{1+o\left(1\right)\} \\
&= q^{2t-r}\left(\frac{q^t+q^{r-t}-2}{q^t}\right)^k\{1+o\left(1\right)\} \quad (k\to\infty).
\end{aligned}$$

This proves Lemma 2.5, our main correlation bound. □

Continuing the quest for a lower threshold, we compare the means of X, the variable of interest, and Y, the maximum number of disjoint collections of unshattered t-words. Denoting the number of overlapping *pairs* of unshattered t-words by Z, we have that

$$Y \le X \le Y + Z,$$

so that

$$\mathbb{E}(X) \le \mathbb{E}(Y) + \mathbb{E}(Z).$$

Now, Fact 10.1 in [9] can be adapted to our case as follows:

Lemma 2.6. *Let m denote the median of Y. Then*

$$|\mathbb{E}\left(Y\right) - m| \le 40\sqrt{kt\mathbb{E}\left(Y\right)},$$

and so, by (1),

$$
\begin{aligned}
\mathbb{P}(X=0)=\mathbb{P}(Y=0) &\leq 2e^{-\frac{m}{4kt}} \\
&\leq 2e^{-\frac{1}{4kt}\{\mathbb{E}(Y)-40\sqrt{kt\mathbb{E}(Y)}\}} \\
&\leq 2e^{-\frac{1}{4kt}\{\mathbb{E}(X)-\mathbb{E}(Z)-40\sqrt{kt\mathbb{E}(X)}\}}. \qquad (4)
\end{aligned}
$$

The key issue is thus to find conditions under which $\mathbb{E}(Z) \to 0$. By Lemma 2.5,

$$
\begin{aligned}
\mathbb{E}(Z) &= \sum_{\Gamma\cap\Delta\neq\emptyset} \mathbb{P}(I_\Gamma I_\Delta = 1) \\
&\leq \sum_{j=1}^{\binom{n}{t}} \sum_{r=1}^{t-1} \binom{t}{r}\binom{n-t}{t-r} q^{2t-r} \left(\frac{q^t+q^{r-t}-2}{q^t}\right)^k (1+o(1)) \\
&\leq K \sum_{r=1}^{t-1} n^{2t-r} \left(\frac{q^t+q^{r-t}-2}{q^t}\right)^k \qquad (5)
\end{aligned}
$$

for some constant $K = K_{t,q}$. The rth term in (5) tends to zero provided that

$$
k \geq \frac{(2t-r)\lg n + \omega(n)}{\lg\left(\frac{q^t}{q^t+q^{r-t}-2}\right)}, \qquad (6)
$$

with $\omega(n) \to \infty$ being arbitrary. The next two lemmas enable us to determine when (6) holds for all r.

Lemma 2.7. *The function $f : [2, t-1] \to \mathbb{R}$ defined by $f(r) = \frac{q^{r-1}-1}{r-1}, q \geq 2$ is monotonically increasing.*

Proof. Define $g(a) = \frac{q^a-1}{a}$; for $1 \leq a \leq t-2$. We have

$$
g'(a) = \frac{aq^a \log(q) - (q^a-1)}{a^2}
$$

The result follows since $aq^a \log q - (q^a-1) = q^a(a\log q - 1) + 1 \geq 2(\log 2 - 1) + 1 = 2\log 2 - 1 > 0$. □

Lemma 2.8. *The constant $\frac{2t-r}{\lg\left(\frac{q^t}{q^t+q^{r-t}-2}\right)}$ indicated by (6) is largest when $r = 1$.*

Proof. We prove that $\frac{2t-r}{\lg\left(\frac{q^t}{q^t+q^{r-t}-2}\right)} \leq \frac{2t-1}{\lg\left(\frac{q^t}{q^t+q^{1-t}-2}\right)}$ for integers $q \geq 2$, $t \geq 3$, and $r \in [2, t-1]$. This occurs if and only if

$$\left(\frac{q^t}{q^t+q^{1-t}-2}\right)^{2t-r} \leq \left(\frac{q^t}{q^t+q^{r-t}-2}\right)^{2t-1},$$

or

$$\left(1+\frac{q^{1-t}\left(q^{r-1}-1\right)}{q^t+q^{1-t}-2}\right)^{2t-1}\left(1+q^{1-2t}-2q^{-t}\right)^{r-1} \leq 1.$$

Since $1+x \leq e^x$, it suffices to show that:

$$\exp\left\{\left(\frac{(2t-1)\,q^{1-t}\left(q^{r-1}-1\right)}{q^t+q^{1-t}-2}\right)+(r-1)\left(q^{1-2t}-2q^{-t}\right)\right\} \leq 1,$$

or

$$(2t-1)\left(\frac{q^{r-1}-1}{r-1}\right) < 2q^{t-1}-1+\frac{4}{q^t}-\frac{1}{q^{2t-1}}-\frac{4}{q}.$$

Because $\frac{4}{q^t} > 0$ and $1+\frac{1}{q^{2t-1}}+\frac{4}{q} \leq 1+\frac{1}{32}+2 \leq 4$, it then suffices to show that

$$(2t-1)\left(\frac{q^{r-1}-1}{r-1}\right) < 2q^{t-1}-4,$$

or, by Lemma 2.7, that

$$\frac{2t-1}{t-2}(q^{t-2}-1) \leq 2(q^{t-1}-2). \tag{7}$$

Now (7) may be verified to be true for $t \geq 4; q \geq 2$ and for $t = 3, q \geq 3$. The remaining case, $t = 3; q = 2$ can be checked by verifying the statement of Lemma 2.8 directly. This completes the proof. □

By (5) and Lemma 2.8,

$$\begin{aligned}\mathbb{E}(Z) &\leq K_{t,q} n^{2t-1}\left(\frac{q^t+q^{1-t}-2}{q^t}\right)^k \\ &\to 0 \quad \text{if } k \geq \frac{(2t-1)\lg n}{\lg\left(\frac{q^t}{q^t+q^{1-t}-2}\right)}(1+o(1));\end{aligned} \tag{8}$$

the next lemma verifies the rather critical fact that this occurs for k's that are smaller than the lower threshold we hope to exhibit.

Lemma 2.9.
$$\frac{(2t-1)}{\lg\left(\frac{q^t}{q^t+q^{1-t}-2}\right)} < \frac{t}{\lg\left(\frac{q^t}{q^t-1}\right)}$$

Proof. The claim is equivalent to

$$\left(\frac{q^t}{q^t-1}\right)^{2t-1} < \left(\frac{q^t}{q^t+q^{1-t}-2}\right)^t,$$

or to
$$\frac{q^{2t}}{q^{2t}-2q^t+1}\left(\frac{q^t-1}{q^t}\right)^{1/t} < \frac{q^t}{q^t+q^{1-t}-2}.$$
Using the inequalities $1-x \le e^{-x}$ and $e^{-x} \le 1/(1+x)$, we see that

$$\left(\frac{q^t-1}{q^t}\right)^{1/t} \le \exp\{-1/(tq^t)\} \le \frac{tq^t}{1+tq^t},$$

so that it would suffice to show, on simplification, that

$$q^t\left\{t(q-1)+2\right\} < q^{2t}+1,$$

which is true since $q^t \ge t(q-1)+2; q,t \ge 2$. □

We are now ready to state the main result of this section.

Theorem 2.10. *Consider a $k \times n$ array with entries that are uniformly and independently selected from $\{0,1,\ldots,q-1\}$. Then if for large enough $A = A_{t,q}$*
$$k \le \frac{t\lg n - \omega(n)}{\lg\frac{q^t}{q^t-1}},\ \ \omega(n) \ge A\lg\lg n,$$
then the probability that the array is a $(t,q,n,k,1)$-covering array tends to zero as $n \to \infty$.

Proof. Equation (4) reveals that the array will be t-covering with low probability whenever $\mathbb{E}(Z) \to 0$ and $\mathbb{E}(X)/kt \to \infty$. Since

$$\mathbb{E}(X) \ge \binom{n}{t}\left(\frac{q^t-1}{q^t}\right)^k \ge \frac{n^t}{t!}\left(\frac{q^t-1}{q^t}\right)^k(1+o(1)),$$

we have that $\mathbb{E}(X)/kt \to \infty$ provided that

$$k \leq \frac{t\lg n - \omega(n)}{\lg\left(\frac{q^t}{q^t-1}\right)}$$

with $\omega(n) \geq A\lg\lg n$. Thus (8) and Lemma 2.9 reveal that $\mathbb{P}(X=0) \to 0$ if

$$\frac{(2t-1)\lg n}{\lg\left(\frac{q^t}{q^t+q^{1-t}-2}\right)}(1+o(1)) \leq k \leq \frac{t\lg n - \omega(n)}{\lg\left(\frac{q^t}{q^t-1}\right)},\ \ \omega(n) \geq A\lg\lg n.$$

The full conclusion of the theorem follows by monotonicity, in k, of $\mathbb{P}(X=0)$. $\square$

To seal the connection between covering arrays and shattering multisets, we restate Theorem 2.10 as follows:

Theorem 2.11. *Consider k multisets $\mathcal{A} = \{A_1, \ldots, A_k\}$ of $[n]$ as follows:*

(i) Each element of $[n]$ is represented in A_i at most $q-1$ times; $q \geq 2$, $1 \leq i \leq k$, and

(ii) The ensemble $\mathcal{A}$ is randomly generated by choosing the multiplicity of each element j in multiset A_i $(1 \leq j \leq n; 1 \leq i \leq k\}$ independently and uniformly from the set $\{0, 1, \ldots, q-1\}$.

Then the collection $\mathcal{A}$ fails to shatter all multisets of t elements, each element repeated $q-1$ times, with high probability if $k \leq (t\lg n - \omega(n))/\lg\frac{q^t}{q^t-1}$, $\omega(n) \geq A\lg\lg n$.

Together with Theorem 2.1, Theorems 2.10 and 2.11 show that the gap between the lower and upper thresholds is rather small; actually this gap arises as an artifact of the Talagrand inequality – and may in fact be artificial. Finally, observe that we have actually uncovered a threshold for the VC dimension of random multiset arrays. For example, if $q = 2$ then sets of size 3 are fully shattered with high probability (w.h.p.) at the level $3\lg n/\lg(8/7) \approx 15.57\lg n$. Thus the VC dimension is 4 or more. But few sets of size 4 are shattered at this level; they are all shattered w.h.p. when the number of rows are of magnitude $4\lg n/\lg(16/15) \approx 42.96\lg n$. In between these levels, the VC dimension of the set system is thus *equal to* 4 w.h.p.

3 Shattering Permutations

For permutations, we use a model analogous to the one used for words in the previous section. Let $\mathcal{S}$ be a randomly generated set of k permutations $\pi_1, \ldots, \pi_k \in S_n$, with each chosen independently with probability $1/n!$ As before, we can represent $\mathcal{S}$ as an array, and we will continue to use *rows* to refer to the elements of $\mathcal{S}$ and *columns* to refer to the positions within each element of $\mathcal{S}$.

Shattering permutations is conceptually similar to shattering words. Let i_1, i_2, i_3 be any 3 elements of $[n]$ and let S^* be the set consisting of the k triples formed by intersecting the i_1th, i_2th, and i_3th columns with the k rows of S. Then, S shatters the triple (i_1, i_2, i_3) (or the *positions* (i_1, i_2, i_3)) if $\rho \in S^*$ up to order-isomorphism for each $\rho \in S_3$.

In this section, we show that the threshold function for the property "shatters all 3-triples" under our model is $k_0(n) = \frac{3}{\lg\left(\frac{6}{5}\right)} \lg(n)$ modulo a small gap. We use the same approach as before, again using Markov's Inequality for the upper threshold and Talagrand's Inequality for the lower threshold. (The analysis becomes intractable for higher values of t, which is the size of the tuple we wish to shatter, and we thus restrict to $t = 3$ in this paper.)

Theorem 3.1. *If $k \geq \frac{3}{\lg\left(\frac{6}{5}\right)} \lg(n)(1 + o(1))$, then all triples are shattered almost surely by $\mathcal{S}$.*

Proof. Let X be the number of unshattered triples. By Markov's Inequality, we have:

$$\mathbb{P}(X \geq 1) \leq \mathbb{E}(X) \leq \binom{n}{3} 6 \left(\frac{5}{6}\right)^k \leq \frac{n^3}{3!} 6 \left(\frac{5}{6}\right)^k \to 0$$

provided that $k \geq \frac{3 \lg n + \omega(n)}{\lg\left(\frac{6}{5}\right)}$. This proves the result. $\square$

Now, for the lower threshold. Define Y as the maximum number of non-overlapping sets of unshattered triples of positions. Y is clearly 1-Lipschitz with certification function $f(s) = 3ks$, and as before

$$Y \leq X \leq Y + Z,$$

with Z defined as the number of pairs of overlapping unshattered triples of positions. The correlation between overlapping triples in the same row is crucial in understanding the quantity

$$\mathbb{E}(Z) = \sum_{\Gamma\cap\Delta\neq\emptyset} \mathbb{P}(I_\Gamma I_\Delta = 1) = \sum_{|\Gamma\cap\Delta|=1} \mathbb{P}(I_\Gamma I_\Delta = 1) + \sum_{|\Gamma\cap\Delta|=2} \mathbb{P}(I_\Gamma I_\Delta = 1),$$

where Γ and Δ range over the set of distinct non-disjoint sets of 3 columns, and the indicator random variable I_Γ is defined as being 1 if Γ is missing at least one 3-permutation and 0 otherwise. Unlike the case of words, the situation here is more complex. First, given an overlap size we can no longer consider just two cases, since, e.g., two patterns ijk and abc may correspond to overlapping last and first columns in Γ, Δ respectively, but k may equal 3 and a might be 1. Second, we can no longer assume without loss of generality, as we did with words, that the overlap occurred in the last r columns of Γ, and that the rest of Δ was entirely to the right of that overlap. In other words, correlations depend not just on the magnitude of the overlap, but its nature as well. With this in mind, let A_m be the event that exactly m 3-permutations are missing from $\Gamma, 1 \le m \le 5$, let $B_{abc} \subseteq A_1$ be the event that the 3-permutation abc is the only permutation missing from Γ, and let C_{ijk} be the event that ijk is missing from Δ. Then, we have:

$$\begin{aligned}
\mathbb{P}\left(I_\Gamma I_\Delta = 1\right) &= \mathbb{P}\left(I_\Gamma = 1\right)\mathbb{P}\left(I_\Delta = 1 | I_\Gamma = 1\right) \\
&= \mathbb{P}\left(I_\Gamma = 1\right)\mathbb{P}\left(I_\Delta = 1 | A_1 \cup A_2 \cup \cdots \cup A_5\right) \\
&= \mathbb{P}\left(I_\Gamma = 1\right)\left[\frac{\mathbb{P}\left(I_\Delta = 1 \cap A_1\right) + \cdots + \mathbb{P}\left(I_\Delta = 1 \cap A_5\right)}{\mathbb{P}\left(A_1 \cup A_2 \cup \ldots \cup A_5\right))}\right] \\
&\le \mathbb{P}\left(I_\Gamma = 1\right)\left[\mathbb{P}\left(I_\Delta = 1 | A_1\right) + \left(\frac{\mathbb{P}\left(A_2\right) + \cdots + \mathbb{P}\left(A_5\right)}{\mathbb{P}\left(A_1\right) + \mathbb{P}(A_2) + \ldots \mathbb{P}(A_5)}\right)\right] \\
&= \mathbb{P}\left(I_\Gamma = 1\right)\left[\mathbb{P}\left(I_\Delta = 1 | A_1\right) + \frac{15 \cdot (4/6)^k}{6 \cdot (5/6)^k - 15 \cdot (4/6)^k}\right] \\
&= \mathbb{P}\left(I_\Gamma = 1\right) \cdot \\
&\quad \left[\mathbb{P}\left(I_\Delta = 1 | B_{123} \cup B_{132} \cup B_{213} \cup B_{231} \cup B_{312} \cup B_{321}\right) + O((4/5)^k)\right] \\
&= \mathbb{P}\left(I_\Gamma = 1\right)\left[\frac{\mathbb{P}\left(I_\Delta = 1 \cap B_{123}\right) + \ldots + \mathbb{P}\left(I_\Delta = 1 \cap B_{321}\right)}{\mathbb{P}\left(B_{123}\right) + \mathbb{P}\left(B_{132}\right) + \cdots + \mathbb{P}\left(B_{321}\right)} + O((4/5)^k)\right] \\
&\le \mathbb{P}\left(I_\Gamma = 1\right)\left(\mathbb{P}\left(I_\Delta = 1 | B_{123}\right) + \ldots + \mathbb{P}\left(I_\Delta = 1 | B_{321}\right)) + O((4/5)^k)\right) \\
&\le 6\left(\frac{5}{6}\right)^k \left(\mathbb{P}\left(C_{123} | B_{123}\right) + \ldots + \mathbb{P}\left(C_{321} | B_{321}\right)\right) + O((2/3)^k).
\end{aligned} \tag{9}$$

Accurate estimation of the quantities $\mathbb{P}\left(C_{ijk}|B_{abc}\right)$ will thus be critical.

Let Γ and Δ be distinct non-disjoint sets of 3 columns. Let D_{abc} and F_{ijk} be the events that abc appears in a fixed row of Γ and ijk appears in the same row of Δ. Then $\mathbb{P}(D_{abc} \cap F_{ijk}) = A/120$ or $\mathbb{P}(D_{abc} \cap F_{ijk}) = B/24$ in the "one-overlap" and "two-overlap" cases respectively, where A and B are the number of ways the two patterns can co-exist among the five or four numbers in the two sets of columns. These numbers can, without loss of generality can be taken to be 1, 2, 3, 4, and 5; or 1, 2, 3, and 4 in the one-overlap and two-overlap cases respectively. Now the probability distribution of the components of Γ conditional on B_{abc} can be obtained, as in Section 2, by randomly selecting five rows; placing one pattern other than abc in these rows; and randomly choosing a pattern other than abc to appear in the other rows. For simplicity, however, we will assume that *each* of the k rows in Γ is equally likely to be chosen to be one of the non-abc patterns; it can be shown that using this slightly incorrect[1] conditional distribution leads to no change in our final conclusion. We have

$$\begin{aligned}
\mathbb{P}\left(C_{ijk}|B_{abc}\right) &= \left(\frac{\sum_{uvw\neq abc}\mathbb{P}(D_{uvw}\cap F_{ijk}^C)}{\sum_{uvw\neq abc}\mathbb{P}(D_{uvw})}\right)^k \\
&= \left(\sum_{uvw\neq abc}\frac{6}{5}\cdot\mathbb{P}(D_{uvw}\cap F_{ijk}^C)\right)^k. \qquad (10)
\end{aligned}$$

Lemma 3.2. *Assume that $|\Gamma\cap\Delta| = 1$ and let γ,δ refer to the index of the overlapping position in the two sets of columns. Then we have the probabilities in Table 1.*

Proof. We will exhibit just two of the calculations using two different proofs; the other calculations may be performed similarly using either of these methods. It is easier to calculate the probabilities $\mathbb{P}(D_{uvw}\cap F_{ijk}|B_{abc})$ instead of $\mathbb{P}(D_{uvw}\cap F_{ijk}^C|B_{abc})$. First suppose that $uvw = 321$, $ijk = 132$ and (without loss of generality) $abc = 123$. Let the five positions spanned by (Γ,Δ) be as follows (this is the $\gamma = 2; \delta = 1$ case):

Γ	3	2	1		
Δ		1		3	2

[1] since this yields a non-zero probability of there being four or fewer patterns in Γ.

(γ,δ)	$\mathbb{P}(D_{uvw}\cap F^C_{ijk}\mid B_{abc})$
(1,1)	$\frac{14}{100}$
(1,2)	$\frac{17}{100}$
(1,3)	$\frac{19}{100}$
(2,2)	$\frac{16}{100}$
(2,3)	$\frac{17}{100}$
(3,3)	$\frac{14}{100}$

Table 1: $\mathbb{P}(D_{uvw}\cap F^C_{ijk}|B_{abc})$ for Overlap 1

The first thing to observe is that the relative positions of the non-overlapping indices amongst the five positions are irrelevant. Denoting the numbers in the five positions, from smallest to largest, by 1, 2, 3, 4, and 5, we see that the arrangement above is fulfilled by the permutations 32154, 42153, and 52143; thus $\mathbb{P}(D_{321}\cap F_{132}|B_{123}) = \frac{6}{5}\cdot\frac{3}{120} = \frac{3}{100}$, whence $\mathbb{P}(D_{321}\cap F^C_{132}|B_{123}) = \frac{1}{5} - \frac{3}{100} = \frac{17}{100}$. This proves the validity of the second entry in Table 1. Let us next verify the fifth entry using another method. Assuming that the alignment of Γ, Δ is as below

$$\begin{array}{cccccc} \Gamma & 1 & 3 & & 2 & \\ \Delta & & & 2 & 3 & 1 \end{array}$$

we choose the common element, x, and must then choose one element larger than, and one smaller than x in the Γ columns, and two elements smaller than x for the columns in Δ. This yields, e.g.,

$$\begin{aligned} \mathbb{P}(D_{132}\cap F_{231}) &\approx \sum_{x=3}^{n-1}\left(\frac{1}{n-3}\right)\left(\frac{x-1}{n-1}\right)\left(\frac{n-x}{n-2}\right)\left(\frac{1}{2}\right)\left(\frac{x-1}{n-1}\right)\left(\frac{x-2}{n-2}\right) \\ &\approx \frac{1}{2}\int_0^1 x^3(1-x)dx \\ &= \frac{1}{40}, \end{aligned}$$

so that $\mathbb{P}(D_{132}\cap F^C_{231}|B_{123}) = \frac{17}{100}$, as asserted; the approximations in the above equation array actually give an exact answer due to combinatorial considerations. $\square$

Returning to (9), we first address the contribution of the $O(2/3)^k$ term. Summing this quantity over all choices of Γ, Δ, we see that the net contribu-

tion of terms corresponding to two or more permutations being absent in Γ is negligible provided that

$$An^5\left(\frac{2}{3}\right)^k \to 0$$

for some constant A. This occurs if $k \geq (5+o(1))\mathrm{lg}n/\mathrm{lg}(1.5)$, or if $k \geq 8.55\mathrm{lg}n$.

We now turn to the 36 terms in the first part of the last line in (9), each term of which may be calculated using (10). Rather than work each term separately, we find the worst (largest) term and use it as a bound for the others. However, doing one calculation in detail will be instructive and seen to be quite general. Assume that $abc = 123$ and $ijk = 231$. The indices uvw now vary among $132, 213, 231, 312$, and 321. Since $|\Gamma \cap \Delta| = 1$, we see that no matter how the overlap occurs in the five columns that determine $\Gamma \cup \Delta$, the indices $\gamma\delta$ consist of two dgs, two egs and one fg – where g is the same no matter how we vary uvw, and $\{d, e, f\} = \{1, 2, 3\}$. Table 1 now reveals that with $g = 1$, the three relevant numbers are 14/100, 17/100, and 19/100; and that these triads are 17/100, 16/100, and 17/100 ($g = 2$); and 19/100, 17/100, and 14/100 ($g = 3$). Maximizing over these possibilities, we see that

$$\mathbb{P}(C_{ijk}|B_{abc}) \leq (0.86)^k,$$

and thus for a constant B, the contribution of the single overlap case to the correlation in (9) and thus to the value of $\mathbb{E}(Z)$ is $O(n^5(5/6 \cdot (0.86))^k$, which tends to zero if

$$k \geq 10.41\mathrm{lg}n\left(\geq \frac{5}{\mathrm{lg}1.395}\mathrm{lg}n\right).$$

NOTE: Observe that a “global maximization” calculation with two 19/100’s and three 17/100’s would have yielded $\mathbb{E}(Z) \to 0$ if $n^5(5/6 \cdot (0.89))^k \to 0$, or if $k \geq 11.59\mathrm{lg}n$ – invalidating our proposed method of proof, since the putative lower threshold is at $3\mathrm{lg}n/\mathrm{lg}(1.2) = 11.405\mathrm{lg}n$.

Consider next the two-overlap case. Equation (10) remains unchanged, but the conditional probabilities $\mathbb{P}(D_{uvw} \cap F^C_{ijk}|B_{abc})$ have a denominator of 20. Given the four entries in $\Gamma \cup \Delta$ in the two-overlap case, the components of the pattern uvw may be chosen in 4 ways, and it remains to calculate how many of these are also consistent with F_{ijk}. There are three cases. If the

components of the two columns in the overlap are *identical* as, e.g., in

$$\begin{array}{cccc} 3 & 2 & 1 & \\ & 2 & 1 & 3 \end{array}$$

the entries in the four positions may appear in two forms, in this case 3214 or 4213. If the components of the two columns are *consistent* as, e.g., in

$$\begin{array}{cccc} 2 & 1 & 3 & \\ & 2 & 3 & 1 \end{array}$$

then there is only one possible arrangement, in this case 3241. Finally, if the components are *inconsistent*, consisting of one monotone increasing and one monotone decreasing pattern, e.g.,

$$\begin{array}{cccc} 1 & 3 & & 2 \\ 3 & & 2 & 1 \end{array}$$

then there is clearly no arrangement. We thus have $\mathbb{P}(D_{uvw} \cap F_{ijk})$ equalling 2/24, 1/24, or 0 in these three cases, or $\mathbb{P}(D_{uvw} \cap F_{ijk}^C)$=2/24, 3/24, or 4/24 respectively. Now given a taboo index abc, a target pattern ijk and any two overlap positions, we upper bound as follows: There can be at most three component arrangements $uvw \neq abc$ inconsistent with ijk. Of the remaining two arrangements uvw, the worst case is when both are consistent with the overlapping positions. This possibility is actually realized with $ijk = 132, abc = 123$, and the overlap occurring in the first and third spots of Δ. (9) and (10) thus yield for some constant C,

$$\mathbb{P}(I_\Gamma I_\Delta = 1) \leq C \cdot \left(\frac{5}{6}\right)^k \cdot \left(\frac{6}{5}\left\{\frac{12}{24} + \frac{6}{24}\right\}\right)^k = C \cdot (0.75)^k$$

and thus the contribution of the double overlap case to $\mathbb{E}(Z)$ is of magnitude $n^4(0.75)^k$, which tends to zero if $k \geq 9.64\lg n$. We are now ready to prove the main result of this section:

Theorem 3.3. *Suppose we choose k permutations randomly, uniformly, and with replacement from S_n, then the probability that they shatter all 3-permutations in any three positions $i_1 < i_2 < i_3$ tends to zero as $n \to \infty$ provided that $k \leq \frac{3\lg n - \omega(n)}{\lg 1.2}$ where $\omega(n) \geq E\lg\lg n$ for some constant E.*

Proof. By Talagrand's inequality used as in Section 2, with $m = \mathrm{Med}(Y)$, we see that

$$\mathbb{P}(Y = 0) \le 2e^{-\frac{m}{12k}},$$

and thus, since

$$|\mathbb{E}(Y) - m| \le 40\sqrt{3k\mathbb{E}(Y)},$$

it follows that

$$\mathbb{P}(X = 0) = \mathbb{P}(Y = 0) \le 2\exp\left(-\frac{1}{12k}(\mathbb{E}(X) - \mathbb{E}(Z) - 40\sqrt{3k\mathbb{E}(X)})\right).$$

As in Theorem 2.10, we thus need to verify when $\mathbb{E}(Z) \to 0$, and $\mathbb{E}(X)/k \to \infty$. We have already seen that $\mathbb{E}(Z) \to 0$ when $k \ge 10.41\mathrm{lg}n$. Finally, we have for constants D, E

$$\mathbb{E}(X) \ge Dn^3\left(\frac{5}{6}\right)^k \to \infty$$

if

$$k \le \frac{3\mathrm{lg}n - \omega(n)}{\mathrm{lg}(1.2)},$$

with $\omega(n) \ge E\mathrm{lglg}n$. The full conclusion of the theorem, namely that $\mathbb{P}(X = 0) \to 0$ for $k \le 10.41\mathrm{lg}n$, follows by monotonicity. □

4 Open Questions

1. Can the results in this paper be of value in actually improving on minimum known sizes of covering arrays for small values of n as found, e.g., in websites maintained by Charlie Colbourn and the NIST? To answer this question, one would first need to derive exact and messy upper and lower bounds for the probability that a random array produces a covering, devoid of $(1 + o(1))$ terms. This can then be translated into a statement about the expected number of random arrays needed until a covering array can be found. Would this approach be computationally feasible, and for which values of q, t could one hope for an improvement?

2. Can threshold results be obtained for probability models other than the ones adopted by us?

3. Can the results of Section 3 be extended to all vaues of t? Is the easy-to-prove upper threshold of $\frac{t \lg n}{\lg(t!/(t!-1))}$ close to the lower threshold?

5 Acknowledgments

The research of all three authors was supported by NSF Grant 1004624.

References

[1] N. Alon and J. Spencer (1992). *The Probabilistic Method*. Wiley, New York.

[2] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. Warmuth (1989). "Learnability and the Vapnik-Chervonenkis dimension," *J. Assoc. Computing Machinery* **36**, 929–965.

[3] M. Bona (2012). *Combinatorics of Permutations.* CRC Press, Boca Raton.

[4] J. Cibulka and J. Kyncl (2012). "Tight bounds on the maximum size of a set of permutations with bounded VC-dimension," *J. Combin. Theory Ser. A* **119**, 1461–1478.

[5] C. Colbourn (2004). "Combinatorial aspects of covering arrays," *Le Matematiche (Catania)* **58**, 121-167.

[6] R. Dudley (1978). "Central limit theorems for empirical measures," *Ann. Probab.* **6**, 899–929.

[7] A. Godbole, D. Skipper, and R. Sunley (1996). "t-Covering arrays: upper bounds and Poisson approximations," *Combinatorics, Probability and Computing* **5**, 105-117.

[8] D. Knuth (1973). *The art of computer programming. Volume 3: Sorting and Searching.* Addison-Wesley Publishing Co., Reading, Massachusetts.

[9] M. Molloy and B. Reed (2001). *Graph Colouring and the Probabilistic Method*. Springer-Verlag, Berlin.

[10] A. Marcus and G. Tardos (2004). "Excluded permutation matrices and the Stanley-Wilf conjecture," *J. Combin. Theory Ser. A* **107**, (2004), 153-160.

[11] N. Sloane (1993). "Covering arrays and intersecting codes," *J. Combin. Designs* **1**, 51-63.

[12] V. Vapnik, (1999). *The Nature of Statistical Learning Theory.* Springer Verlag, New York.

[13] V. Vapnik and A. Chervonenkis (1971). "On the uniform convergence of relative frequencies of events to their probabilities," *Theory Probab. Appl.* **16**, 264–280.